Identity, boundaries, and categories: Socio-geographical guidance on standards and infrastructure development

Mark Parsons (1), Matthew Mayernik (2), and Peter Fox (1)

(1) Tetherless World Constellation, Rensselaer Polytechnic Institute, Troy, NY, USA, (2) National Center for Atmospheric Research, Boulder, CO, USA

The persistent, unambiguous identification of data objects is now recognized as a necessary and central component of modern data infrastructure, especially in an interdisciplinary context. Registered persistent identifiers work to allow machines and humans to understand which digital object is in question (identity), what it is (category or type), and where it is (location). Each of these questions is surprisingly fraught and complex. This presentation begins to explore how a socio-geographical perspective can help us begin to address that complexity and answer those seemingly basic questions.

We explore how issues of data identity are very dependent on the context in which the data are perceived. Using Susan Leigh Star's concept of "boundary objects" — objects that are part of multiple social worlds and facilitate communication between them — we examine how data and even their identifiers can be adaptable to different interpretations and categorizations. The idea with a boundary object is to reduce local uncertainty without damaging cooperation, where representations of data and identity support multiple interpretations and purposes across intersecting worlds without a shared consensus. We further draw from human geographic concepts of borderlands and feminist ideas of intersectionality to illustrate new ways of considering how data is managed, described, and presented.

We close with examples of how these socio-geographic considerations can pragmatically guide decisions in the development of standards and data infrastructure ranging from specific details around metadata management and data formats to more holistic considerations, such as community development and standards uptake.